FILTER DESIGN FOR DELAY-BASED ANONYMOUS COMMUNICATIONS

Simon Oya[*]   *Fernando Pérez-González*[*]   *Carmela Troncoso*[†]

[*]Signal Theory and Communications Dept., University of Vigo
[†]The IMDEA Software Institute

ABSTRACT

In this work, we address the problem of designing delay-based anonymous communication systems. We consider a timed mix where an eavesdropper wants to learn the communication pattern of the users, and study how the mix must delay the messages so as to increase the adversary's estimation error. We show the connection between this problem and a MIMO system where we want to design the coloring filter that worsens the adversary's estimation of the MIMO channel matrix. We obtain theoretical solutions for the optimal filter against short-term and long-term adversaries, evaluate them with experiments, and show how some properties of filters can be used in the implementation of timed mixes. This opens the door to the application of previously known filter design techniques to anonymous communication systems.

Index Terms— Anonymity, filter design, timed mixes, optimization

1. INTRODUCTION

Anonymity in communication systems is typically achieved at the expense of delay or communication bandwidth. Mixes [1, 2], the basic building blocks of high-latency anonymous communication systems, are channels that *delay* messages, change their appearance and output them in a random order in batches. This confuses an eavesdropper trying to unveil the path followed by the messages in the network, who is not able to identify the sender of a message leaving the mix with absolute certainty.

It is well known that two delaying mechanisms that cause the same average delay in the communication can achieve different protection against a malicious observer. Since delay is the main resource to generate privacy in delay-based anonymous communication systems, it is of paramount importance to understand how to use it optimally.

The design of the *delya characteristic* of an anonymous communication system has been studied in the literature with different privacy goals in mind [3, 4, 5]. In this work, we take over the work in [5], where the authors obtain the optimal delay characteristic against an eavesdropper with global vision of the network that tries to learn the average number of messages each sender sends to each receiver. We start from the results in that paper and interpret the problem as a filter design problem in a MIMO system, where the delay characteristic is a filter, which allows us to reason about some results obtained in [5]. We also make a clear distinction between two scenarios that give different optimal delay characteristics, and find the optimal characteristic in an scenario where senders send most of their messages to only one of their friends, which is not covered by previous work. Finally, we show how filter properties can help in the implementation of delay functions in a decentralized way, which is of particular interest in practice.

The rest of the document is distributed as follows. Section 2 presents the system model and notation, as well as previous results that are relevant for this work. It also sets up the optimization problem of designing the delay characteristic that maximizes the privacy of the users. We solve this problem in Sect. 3 for different scenarios, one of which was not considered in previous works, and validate our results. Finally, in Sect. 4 we give examples of how some properties of filter design can be used in the implementation of delay-based anonymous communications, and conclude in Sect. 5.

2. PRELIMINARIES

2.1. System Model and Notation

We consider a system where a group of N senders, indexed by $i \in \{1, \cdots, N\}$, send messages to a group of M receivers, indexed by $j \in \{1, \cdots, M\}$, through an anonymous communication channel, which we model as a *timed mix*. The timed mix contains a timer that loops continuously, counting down starting at τ seconds, thus creating laps that are called *rounds*. When a message arrives at the mix, it is assigned a random delay drawn from a probability mass function called *delay characteristic*. When the timer expires, all the messages whose delay is zero have their appearance changed through cryptographic tools and are forwarded to their corresponding
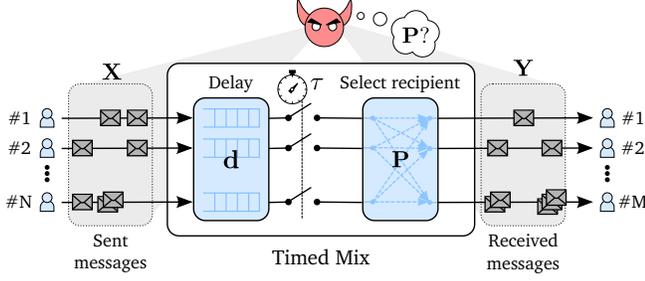
Fig. 1. System model. The adversary observes $\mathbf{X}$ and $\mathbf{Y}$, knows $\mathbf{d}$ and τ, and wants to learn the probabilities in $\mathbf{P}$.

Table 1. Summary of notation

$p_{j,i}$	Prob. user i sends a message to j.	$(\mathbf{P})_{i,j} = p_{j,i}.$	
d_k	Prob. a msg. is delayed k rounds.	$(\mathbf{d})_k = d_k.$	
X_i^r	No. msgs. sent by i in round r.	$(\mathbf{X})_{r,i} = X_i^r.$	
Y_j^r	No. msgs. received by j in round r.	$(\mathbf{Y})_{r,j} = Y_j^r.$	
ν_i	"Sharpness" of i, $\nu_i = \sum_j p_{j,i}^2.$	$(\boldsymbol{\nu})_k = \nu_i.$	

recipients. The messages that remain in the mix have their delay decreased in one unit, and will leave eventually as the timer loops.

We consider an adversary observing the messages arriving and leaving the mix for T seconds, i.e., during a total of $\rho = \lfloor T/\tau \rfloor$ rounds. The adversary can observe all the messages sent by every sender and received by every recipient (i.e., it is *global*), and it is a mere observer of the system (i.e., it is *passive*). One example of such adversary is a malicious Internet Service Provider. The goal of this adversary is to learn the probability that a message sent by i is received by j, denoted by $p_{j,i}$. This represents the percentage of messages from i that go to j on average. We assume that the adversary knows how the mix works (i.e., τ and the delay characteristic) but cannot look inside it. The change of appearance of the messages inside the mix prevents the adversary from performing bit-wise linkability of input and output messages, while the random delay and grouping of messages in rounds prevents timing linkability. This system is depicted in Fig. 1.

We now introduce the statistical model and the notation, which is summarized in Table 1. The random variable that models the number of messages sent by user i in round r is denoted by X_i^r. Each input message is delayed independently and randomly according to the delay characteristic $\mathbf{d} \doteq [d_0, d_1, \cdots, d_{\rho-1}]^T$, where d_k is the probability that the mix delays a message k rounds. The recipient of each message sent by user i is $j \in \{1, \cdots, M\}$ with probability $p_{j,i}$. Finally, the total number of messages that receiver j gets in round r is Y_j^r. Using these variables, we also build the $\rho \times N$ matrix of all input observations $\mathbf{X}$, whose i,r-th element is X_i^r, i.e., $(\mathbf{X})_{r,i} = X_i^r$. Likewise, we build $\mathbf{Y}$ and $\mathbf{P}$ as $(\mathbf{Y})_{r,j} = Y_j^r$ and $(\mathbf{P})_{i,j} = p_{j,i}$.

2.2. Previous Results and Connection with MIMO Communications

Previous works have studied different problems in this same model. Here, we summarize some results, mainly from [5, 6], that are of particular interest for our work. The first result is that $\mathrm{E}\{\mathbf{Y}|\mathbf{X}\} = \mathbf{D} \cdot \mathbf{X} \cdot \mathbf{P}$, where $\mathbf{D}$ is a convolution matrix defined as $(\mathbf{D})_{r,s} = d_{r-s}$ if $r - s \geq 0$, and 0 otherwise.

This means that, on average, the M output processes in the columns of $\mathbf{Y}$ are a linear combination (produced by $\mathbf{P}$) of the N input processes that are the columns of $\mathbf{X}$, convolved with the delay characteristic $\mathbf{d}$. If we define the noise of the outputs $\mathbf{N} \doteq \mathbf{Y} - \mathbf{DXP}$, another result is that the covariance matrix of this noise is $\boldsymbol{\Sigma}_{\mathbf{N}|\mathbf{X}} = \mathrm{diag}\{\mathbf{DX1}_N\} - \mathbf{D} \cdot \mathrm{diag}\{\mathbf{X}\boldsymbol{\nu}\} \cdot \mathbf{D}^T$. In this expression, $\mathbf{1}_N$ is an $N \times 1$ all-ones vector, and $\boldsymbol{\nu}$ is a vector whose i-th entry is $\nu_i \doteq \sum_{j=1}^{M} p_{j,i}^2$. This parameter represents the "sharpness" of the sending behavior of user i. Values $\nu_i \approx 1$ represent a sender that focuses in a single receiver, while $\nu_i \approx 0$ represents a sender that distributes messages evenly among her recipients.

From these results, we can see the $\mathbf{d}$ box in Fig. 1 as a linear filter and $\mathbf{P}$ as a MIMO channel matrix for the averages, and consider that $\mathbf{Y}$ is obtained after adding the noise $\mathbf{N}$ to this average. The problem of designing the delay characteristic $\mathbf{d}$ against an adversary that wants to estimate the sending behavior $\mathbf{P}$ is then equivalent to the problem of designing a *filter* $\mathbf{d}$ against an adversary that wants to estimate the MIMO channel matrix $\mathbf{P}$. We will use this interpretation of $\mathbf{d}$ as a filter below. Note that even though $\mathbf{X}$ and $\mathbf{Y}$ only contain integer numbers (because messages cannot be broken in smaller units), with this interpretation we can disregard the integer constraints because $\mathbf{X}$ and $\mathbf{Y}$ are observations, not parameters to estimate.

2.3. Privacy Metric and Analysis

As mentioned above, the adversary wants to learn the probabilities $\mathbf{P}$, that are sensitive information from the users, after observing $\mathbf{X}$ and $\mathbf{Y}$ and knowing τ and $\mathbf{d}$. Previous works delve into the study of the best linear estimator for $\hat{\mathbf{P}}$ in this scenario and the privacy metric used to assess the success of this adversary. We refer to [5, 6] for a thorough derivation of these results, and just note that the overall mean squared error ξ_T of the adversary can be written as $\xi_T = \mathrm{Tr}\{\mathbf{M}\mathbf{C}_e\mathbf{M}\}$, where $\mathbf{M} \doteq \mathrm{diag}\{[\mu(1), \cdots, \mu(N)]\}$ is a normalizing diagonal matrix and $\mathbf{C}_e$ is the covariance matrix of the adversary's error $\mathbf{C}_e \doteq \mathrm{E}\{\mathbf{E}\mathbf{E}^T\}$, where $\mathbf{E} \doteq \hat{\mathbf{P}} - \mathbf{P}$. This matrix can be written for the best linear estimator of $\mathbf{P}$ as

$$\mathbf{C}_e = \mathrm{E}\{(\mathbf{X}^T\mathbf{D}^T\mathbf{D}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{D}^T\boldsymbol{\Sigma}_{\mathbf{N}|\mathbf{X}}\mathbf{D}\mathbf{X}(\mathbf{X}^T\mathbf{D}^T\mathbf{D}\mathbf{X})^{-1}\}. \tag{1}$$

Using the expression for $\boldsymbol{\Sigma}_{\mathbf{N}|\mathbf{X}}$ above, we have a relation between the overall privacy ξ_T and the delay characteristic $\mathbf{d}$.

3. DESIGN OF THE OPTIMAL DELAY CHARACTERISTIC

Now we study how to design the delay characteristic $\mathbf{d}$ such as to increase the adversary's overall error ξ_T. In this section, we use λ_k to denote the k-th DFT coefficient of the ρ-point DFT of $\mathbf{d}$. We start by discussing some constraints on $\mathbf{d}$ and set up the design problem that gives us the optimal filter $\mathbf{d}_{\mathrm{opt}}$. Then, we study the solution to this problem in two cases: 1) when the adversary observes the system indefinitely ($\rho \rightarrow \infty$) and $\rho \gg N$, and 2) when the adversary observes the system a number of rounds ρ commensurate to N. This distinction, which is not clear in previous works, is important because it leads to different solutions. We call each of these scenarios long-term and short-term adversary/attack, respectively.

3.1. Constraints on the Delay Characteristic

Although it is possible to see the delay characteristic $\mathbf{d}$ as a filter, we must keep in mind that it is actually a probability mass function, and therefore it must follow some constraints:

a) Non-negativity: since the values of this filter are probabilities, it must hold that $d_k \geq 0$ for all k. The consequences of non-negativity constraints in filter design have been discussed in [7, 8, 9]. An immediate consequence is that $\lambda_0 \geq \lambda_k$ for all k. A more complex effect of these constraints, explained in detail in [8], is that it is easier to achieve a filter $\mathbf{d}$ with large attenuation factor in high frequencies than in middle and low frequencies.

b) Normalization: another direct consequence of the fact that $\mathbf{d}$ is a probability mass function is $\sum_{k=0}^{\rho-1} d_k = 1$. This forces the first DFT coefficient to be one, i.e., $\lambda_0 = 1$.

c) Average delay: we want to design a mix that guarantees that the average delay of the messages in the system, measured in rounds, does not exceed some value $\bar{\delta}$, i.e., $\sum_{k=0}^{\rho-1} k \cdot d_k \leq \bar{\delta}$. One of the effects of this constraint is that, of all the filters that give the same magnitude response, we will prefer the minimum phase solution, i.e., the one whose zeros lie inside the unit circle. This comes from the fact that the group delay of the filter and the average delay $\bar{\delta}$ are closely related.

We denote the space of filters $\mathbf{d}$ that follow these constraints by $\mathcal{C}$. This is defined formally as

$$\mathcal{C} = \left\{ \mathbf{d} : \sum_{k=0}^{\rho-1} d_k = 1, \sum_{k=0}^{\rho-1} k \cdot d_k \leq \bar{\delta}, d_k \geq 0 \quad \forall k \right\} . \quad (2)$$

Note that the filters $\mathbf{d} \in \mathcal{C}$ are causal (by definition) and stable. Then, the optimal delay characteristic we are looking for, denoted by $\mathbf{d}_{\mathrm{opt}}$, is the solution to

$$\mathbf{d}_{\mathrm{opt}} = \operatorname*{argmax}_{\mathbf{d} \in \mathcal{C}} \operatorname{Tr}\{\mathbf{M}\mathbf{C}_e\mathbf{M}\} . \quad (3)$$

where $\mathbf{C}_e$ is given in (1). We now study this solution against long-term and short-term adversaries.

3.2. Long-Term Optimal Delay Characteristic

When $\rho \rightarrow \infty$ and $\rho \gg N$, we can approximate the terms $\mathbf{X}^T\mathbf{D}^T\mathbf{D}\mathbf{X}$ and $\mathbf{X}^T\mathbf{D}^T\mathbf{\Sigma}_{\mathbf{N}|\mathbf{X}}\mathbf{D}\mathbf{X}$ in (1) by their expected values and obtain a closed form expression for ξ_T. Let $\gamma_1 \doteq \sum_k d_k^2$ and $\gamma_2 \doteq \sum_r \left(\sum_k d_r d_{r+k}\right)^2$. If we assume the inputs are i.i.d. Poissonian, in [10] we show that when users have several friends with whom they communicate evenly ($\nu_i \approx 0$), then ξ_T grows with $1/\gamma_1$. When users send most of their messages to only one of their friends ($\nu_i \approx 1$), then ξ_T grows with $(\gamma_1 - \gamma_2)/\gamma_1^2$. Therefore, we set

$$\mathbf{d}_{\mathrm{long}} = \begin{cases} \operatorname*{argmax}_{\mathbf{d} \in \mathcal{C}} (1/\gamma_1) & \text{if } \nu_i \approx 0 , \\ \operatorname*{argmax}_{\mathbf{d} \in \mathcal{C}} (\gamma_1 - \gamma_2)/\gamma_1^2 & \text{if } \nu_i \approx 1 . \end{cases} \quad (4)$$

The study of $\mathbf{d}_{\mathrm{long}}$ when $\nu_i \approx 0$, including a close-form expression for it, already appears in [5]. However, it is important to note that when $\nu_i \approx 1$ the solution changes, and this case was not analyzed in [5]. Note that we can also extend this solution to the case where the input samples are not independent by including a coloring filter $\mathbf{g}$ before the adversary observation in the model, and replacing $\mathbf{d}$ by $\mathbf{d} * \mathbf{g}$ in the formulas above.

Now we evaluate these results, comparing the numerical solution $\mathbf{d}_{\mathrm{opt}}$ using (3) and the theoretical one $\mathbf{d}_{\mathrm{long}}$ in (4). We generate $\mathbf{X}$ using real data fed to a timed mix. We use the real data in [5] so that this work is comparable to previous ones (see [5] for a thorough description of the real datasets). We take $N = 100$ users and $\rho = 1\,500$ rounds from the datasets in order to study a case where $\rho \gg N$, and generate $\mathbf{P}$ using a Zipf distribution with 10 friends per sender for $\nu_i \approx 0$ and a single friend for $\nu_i = 1$, with $M = 100$. We compute $\mathbf{d}_{\mathrm{opt}}$ and $\mathbf{d}_{\mathrm{long}}$ for the 3 datasets in [5] and show the average results in Fig. 2.a. We can see that the optimal filter obtained analytically (4) is very close to the numerical one obtained through evaluation of (3) using $\mathbf{X}$.

3.3. Short-Term Optimal Delay Characteristic

We now consider the case where the number of observed rounds ρ is commensurate with the number of senders N. In this case, it is argued in [11] that $\mathbf{X}^T\mathbf{D}^T\mathbf{D}\mathbf{X}$ should be made as close to singular as possible. We can write this as $\mathbf{X}_f^H|\mathbf{\Lambda}_d|^2\mathbf{X}_f$ where $\mathbf{X}_f$ is the $\rho \times N$ matrix of the DFT of the inputs and $\mathbf{\Lambda}_d$ is the diagonal matrix containing the coefficients of the ρ-point DFT of $\mathbf{d}$. To make this matrix close to singular, it makes sense to try to make $\rho - N$ DFT coefficients of $\mathbf{d}$ close to zero. This is easier to achieve for high-frequencies, as a consequence of the positivity constraints [8]. Therefore, we can set[1]

$$\mathbf{d}_{\mathrm{short}} = \operatorname*{argmin}_{\mathbf{d} \in \mathcal{C}} \sum_{k=N/2+1}^{\rho-N/2+1} \lambda_k . \quad (5)$$

---

[1]This is for even N. A small tweak is required for odd N.

(a) **d**, long-term adversary, $\nu_i \approx 0$ (left) and $\nu_i = 1$ (right).

(b) **d** (left) and DFT of **d** (right), short-term adversary, $\nu_i \approx 0$.
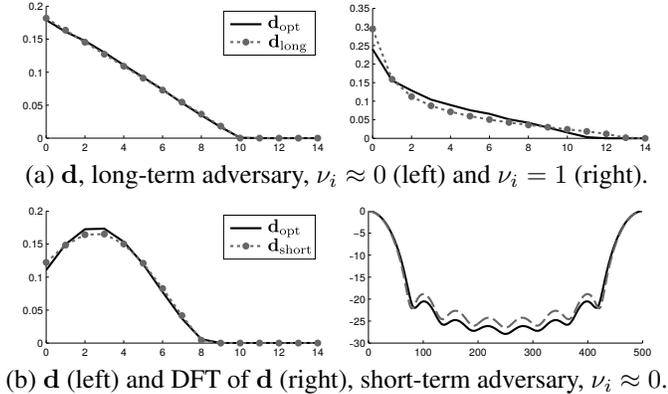
**Fig. 2**. Optimal delay characteristic against a long-term ($\rho = 1\,500$) and short-term ($\rho = 500$) adversary, with $N = 100$.

(a) **d**, single node.

(b) **d** achieved vs. objective.

(c) Frequency response.

(d) Exponential mix as IIR.

(e) Distributed implementation.

Fig. 3. Use of filter properties in anonymous network design.

The results obtained numerically (generating **X** by taking $N = 100$ users and $\rho = 500$ rounds from the real datasets) using (4) and analytically with (5) are close, as shown in Fig. 2.b. From a filter-design perspective, the frequency response in dB (Fig. 2.b, right) confirms that the optimal pool is the low-pass filter that tries to remove information from the $\rho - N = 400$ high-frequency DFT coefficients.

4. APPLICATIONS

In this section, we show how some properties of filters can be used to aid in the implementation of anonymous communication systems. Normally, anonymous communication systems are designed in a non-centralized way, with several devices connected in a network, for scalability and trust reasons [12, 13, 14, 15]. In this case, the delay characteristic of the timed mix is the addition of the individual delays provided by the different components of the network. The task of designing a decentralized delay-based anonymous communication system can be simplified by considering basic properties of filter design. We illustrate the application of some of these properties (direct form implementation, cascade and parallel filters) in the two examples below and leave the study of how these findings can be used to configure and improve real systems like Mixmaster [13] and JAP [15] for future work.

a) Timed mix as a cascade of nodes. We want to implement a time mix in a distributed way using 5 nodes in cascade, to protect $N = 100$ senders against a short-term adversary that observes the inputs and outputs for $\rho = 500$ rounds. The delay introduced by two nodes in cascade properly synchronized with the timer is the convolution of the individual delay characteristics (as in filters in cascade). Therefore, configuring the delay of the nodes as shown in Fig. 3.a, we achieve the overall delay characteristic, result of convolving the 5 individual responses, shown in Fig. 3.b (achieved). We see that this overall delay is close to the optimal one (objective). Fig-

ure 3.c shows the frequency response, where we see that the nodes try to reduce the $\rho - N = 400$ high-frequency DFT coefficients.
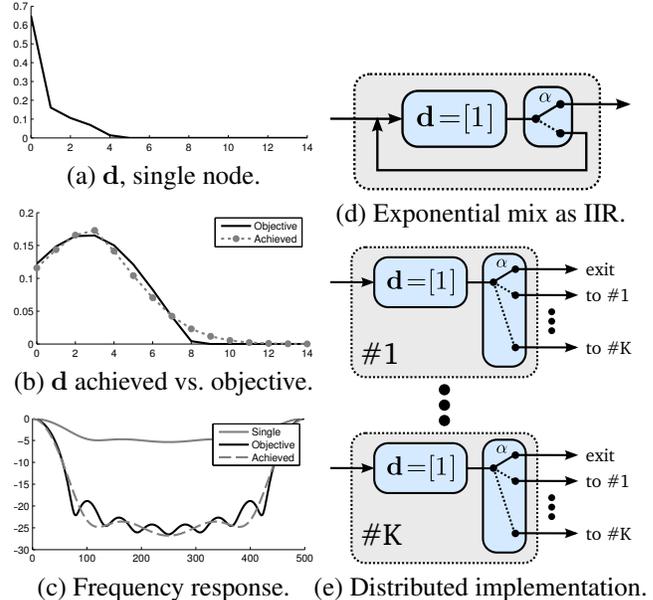
b) Distributed exponential mix. We want to build an exponential mix, i.e., a timed mix with $d_k = \alpha(1 - \alpha)^k$, which is optimal for some privacy metrics [3, 4]. This is equivalent to a first-order IIR filter which we can implement easily in direct form, using a node that only delays messages until the end of a round ($\mathbf{d} = 1$) and a switch that sends messages to an output with probability α, and to the other with probability $1 - \alpha$ (Fig. 3.d). In order to implement this device in a decentralized way, we place several IIR filters in parallel as shown in Fig. 3.e, where a message is forwarded to its recipient with probability α in each round, and fed to a random node with probability $1 - \alpha$. The overall response is the same since this is equivalent to an scenario where we have identical IIR filters in parallel.

5. CONCLUSIONS

In this work, we have studied the problem of designing the delay characteristic of timed mixes against a global adversary that aims at learning the communication patterns of the users. We have obtained theoretical solutions for this problem against a long-term and short-term attack, and have shown through experiments that they are close to the ones obtained numerically. We also showed the connection between designing delay characteristics and filters, and used this connection to implement the timed mixes in a decentralized way as a network of delaying nodes.

6. REFERENCES

[1] D. Chaum, "Untraceable electronic mail, return addresses, and digital pseudonyms," *Commun. of the ACM*, vol. 24, no. 2, pp. 84–90, Feb 1981.

[2] Claudia Diaz and Andrei Serjantov, "Generalising mixes," in *Privacy Enhancing Technologies*, pp. 18–31. 2003.

[3] George Danezis, "The traffic analysis of continuous-time mixes," in *Privacy Enhancing Technologies*, pp. 35–50. 2005.

[4] David Rebollo-Monedero, Javier Parra-Arnau, Jordi Forné, and Claudia Diaz, "Optimizing the design parameters of threshold pool mixes for anonymity and delay," *Computer Networks*, vol. 67, no. 0, pp. 180–200, July 2014.

[5] Simon Oya, Fernando Pérez-González, and Carmela Troncoso, "Design of pool mixes against profiling attacks in real conditions," *IEEE/ACM Transactions on Networking*, vol. 24, no. 6, pp. 3662 – 3675, Dec 2016.

[6] Fernando Pérez-González, Carmela Troncoso, and Simon Oya, "A least squares approach to the static traffic analysis of high-latency anonymous communication systems," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 9, pp. 1341–1355, Sept 2014.

[7] Y Liu and Peter H Bauer, "A new approach to designing high-performance non-negative finite impulse response lowpass filters," in *Signal Processing Systems (SIPS), 2010 IEEE Workshop on*. IEEE, 2010, pp. 444–449.

[8] Y Liu and Peter H Bauer, "Frequency domain limitations in the design of nonnegative impulse response filters," *Signal Processing, IEEE Transactions on*, vol. 58, no. 9, pp. 4535–4546, 2010.

[9] Yuzhe Liu and Peter H Bauer, "Fundamental properties of non-negative impulse response filters," *Circuits and Systems I: Regular Papers, IEEE Transactions on*, vol. 57, no. 6, pp. 1338–1347, 2010.

[10] Simon Oya, Fernando Pérez-González, and Carmela Troncoso, "Technical report tsc/so/24082016: Filter design for delay-based anonymous communications," http://gpsc.uvigo.es/sites/default/files/publications/techrepICASSP2017.pdf.

[11] Simon Oya, Fernando Pérez-González, and Carmela Troncoso, "Technical report for id tnet-2015-00294 "optimal delay characteristic when the number of users is comparable to the number of rounds"," http://gpsc.uvigo.es/sites/default/files/publications/TechRepToN2016.pdf.

[12] G. Danezis, R. Dingledine, and N. Mathewson, "Mixminion: Design of a type iii anonymous remailer protocol," in *IEEE Symposium on Security and Privacy*, 2003, pp. 2–15.

[13] U. Möller, L. Cottrell, P. Palfrader, and L. Sassaman, "Mixmaster Protocol — Version 2," IETF Internet Draft, July 2003.

[14] Ceki Gulcu and Gene Tsudik, "Mixing e-mail with babel," in *Network and Distributed System Security, 1996., Proceedings of the Symposium on*. IEEE, 1996, pp. 2–16.

[15] "Jap: The jap anonymity & privacy homepage," http://www.anon-online.de.